Thermodynamics of DNA Loops with Long-Range Correlated Structural Disorder

C. Vaillant,[1,2] B. Audit,[3] and A. Arnéodo[3]

[1]*Institut Bernouilli, EPFL, 1015 Lausanne, Switzerland*

[2]*Laboratoire Statistique et Génome, 523 place des terrasses de l'Agora, 91000 Evry, France*

[3]*Laboratoire Joliot Curie (CNRS), Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon cedex 07, France*

(Received 9 February 2005; published 1 August 2005)

We study the influence of a structural disorder on the thermodynamical properties of 2D-elastic chains submitted to mechanical/topological constraint as loops. The disorder is introduced via a spontaneous curvature whose distribution along the chain presents either no correlation or long-range correlations (LRC). The equilibrium properties of the one-loop system are derived numerically and analytically for weak disorder. LRC are shown to favor the formation of small loop, larger the LRC, smaller the loop size. We use the mean first passage time formalism to show that the typical short time loop dynamics is superdiffusive in the presence of LRC. Potential biological implications on nucleosome positioning and dynamics in eukaryotic chromatin are discussed.

DOI: 10.1103/PhysRevLett.95.068101 PACS numbers: 87.10.+e, 05.40.−a, 87.14.Gg, 87.15.−v

The dynamics of folding and unfolding of DNA within living cells is of fundamental importance in a host of biological processes ranging from DNA replication to gene regulation [1]. As the basic unit of eukaryotic chromatin organization, the structure and dynamics of nucleosomes have attracted increasing experimental and theoretical interest [2]. High resolution x-ray analyses [3] have provided deep insight into the wrapping of 145 bp of DNA in almost two turns around a histone octamer to form a nucleosome core. Recent experiments have shown that nucleosomes are highly dynamical structures that can be moved along DNA by chromatin remodeling complexes [4] but that can also move autonomously on short DNA segments [5]. Different models have been proposed to account for the nucleosome mobility [6] including the DNA reptation model that involves intranucleosomal loop diffusion [7] and the nucleosome repositioning model via an extranucleosomal loop [8]; both models provide an attractive picture of how a transcribing RNA polymerase can get around nucleosomes without dissociating it completely. Since the discovery of naturally curved DNA [9], several works have investigated the possibility that the DNA sequence may facilitate the nucleosome packaging [10] in the same manner as it can highly promote very small loop formation [11]. Recently, a comparative statistical analysis of eukaryotic DNA bending profiles [12] has revealed that LRC are the signature of the nucleosomal structure and are likely to play a role in the condensation of the nucleosomal string into the 30 nm-chromatin fiber. To which extent sequence-dependent LRC structural disorder does help to regulate the structure and dynamics of chromatin is of fundamental importance in regards to the structural informations that may have been encoded into DNA sequences during evolution. The LRC structural disorder induced by the sequence could favor the formation of small (few hundreds bp) DNA loops and in turn the propensity of eukaryotic DNA to interact with histones to form nucleosomes.

Our aim here is to investigate the influence of LRC structural disorder on the thermodynamical properties of semiflexible chains like DNA when constrained locally to form a loop of size l much smaller than the chain length L. Because of the approximate planarity of nucleosomal DNA loops, one will assume the chains to be confined in a plane and to be free of any twisting deformation. Within the linear elasticity approximation, the local elastic energy variation of a 2D semiflexible chain is:

$$\delta E(s) = \tilde{A}[\dot{\theta}(s) - \dot{\theta}_0(s)]^2/2, \quad (1)$$

where $\tilde{A}$ is the bending stiffness, $\dot{\theta}(s)$ the local curvature, and $\dot{\theta}_0(s)$ the local "spontaneous" curvature of the chain. To model the intrinsic quenched ($T = 0$) disorder, we consider $\dot{\theta}_0(s)$ as the realization of a Gaussian fractional noise of zero mean and variance σ_0^2 and such that the corresponding random walk $\Delta\Theta_0(s, l) = \int_s^{s+l} \dot{\theta}_0(u)du$ exhibits normal fluctuations characterized by:

$$\overline{\Delta\Theta_0}(s, l) = 0, \quad \overline{\Delta\Theta_0^2}(s, l) - \overline{\Delta\Theta_0}^2(s, l) = \sigma_0^2 l^{2H}, \quad (2)$$

where H is called the Hurst exponent [12,13]: when

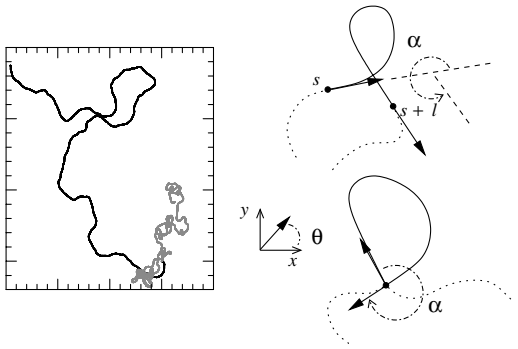

FIG. 1. Left: spontaneous trajectory for two 2D semiflexible chains with uncorrelated ($H = 0.5$, black) and LRC ($H = 0.8$, gray) structural disorder. Right: winding constraint (top) and cyclization constraint (bottom).

$H = 1/2$, one recovers the uncorrelated Gaussian noise; for $H > 1/2$, the distribution of the intrinsic curvature along the chain is LRC. As illustrated in Fig. 1, due to the persistence of the orientation's fluctuations, LRC 2D spontaneous trajectories are more looped than the uncorrelated ones.

To account for the spontaneous formation of a loop of size l, we will consider chains under the following geometrical constraints (Fig. 1): (i) the ''winding'' constraint amounts to keep fixed the variation of the orientation over a length l, $\int_s^{s+l} \dot{\theta}(u)du = \alpha$ and (ii) the ''cyclization'' constraint where in addition the two extremities are held fixed together, $\int_s^{s+l} \cos[\theta(u)]du = \int_s^{s+l} \sin[\theta(u)]du = 0$. Given a chain defined by its spontaneous curvature distribution, we first compute the 1D energy landscape $E(s, l)$ associated to the formation of one loop of length l at the position s. Introducing the constraint via Lagrange multipliers, the equilibrium configuration is obtained by solving the corresponding Euler-Lagrange equations. For the winding constraint, from the equilibrium equations, one gets immediately the shape of the constrained chain and the corresponding energy cost

$$E(s, l) = \tilde{A}[\Delta\Theta_0^2(s, l) - 2\alpha\Delta\Theta_0(s, l) + \alpha^2]/2l. \quad (3)$$

In Fig. 2(a) are shown the energy landscapes for an uncorrelated and a LRC chain; the fluctuations of the latter are much larger than those of the former. In the weak disorder (WD) limit ($\sigma_0^2 \ll 1$), the statistics of the energy landscape is Gaussian; when using Eq. (2), one gets for the mean $\bar{E}(l) = \tilde{A}[\alpha^2/l + \sigma_0^2 l^{2H-1}]/2$ and the variance $\overline{[E(l) - \bar{E}(l)]^2} = \alpha^2\tilde{A}^2\sigma_0^2 l^{2H-2}$. For the cyclization constraint there is no such general analytic derivation of the equilibrium configuration and one has to turn back to numerical computations. As in [14], we have used an iterative scheme to perform numerical computations for several values of α, H, σ_0, and l. In the WD limit, the equilibrium energy fluctuations numerically obtained with the cyclization constraint display Gaussian statistics with the same mean and variance as previously derived with the winding constraint.

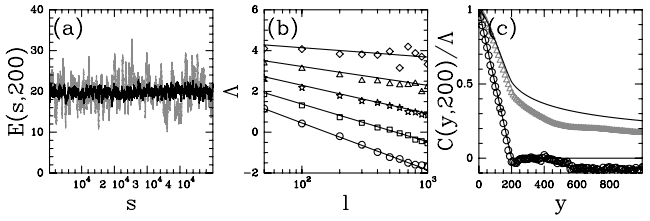

FIG. 2. Single loop of size l in a chain of length $L = 10^5$ under the winding constraint and for $A = 200$, $\sigma_0 = 0.01$, and $\alpha = 2\pi$. (a) Energy landscapes for a $l = 200$ loop along chains with uncorrelated ($H = 0.5$, black) and LRC ($H = 0.8$, gray) disorders. (b) Free energy rms fluctuations $\Lambda(l)$ vs l; the symbols correspond to numerical estimate for different disorders: $H =$ 0.5 (○), 0.6 (□), 0.7 (☆), 0.8 (△), and 0.9 (◇); the solid lines correspond to Eq. (4). (c) Reduced correlation function $C(y, 200)/\Lambda(200)$ vs y; the symbols have the same meaning as in (b); the solid curves correspond to Eq. (5).

At finite temperature, one has to consider the effect of thermal fluctuations which requires us to compute the free energy cost of the loop formation $\beta f(s, l) = \beta E(s, l) - \Delta S(s, l)$, where $\beta = 1/k_B T$. Under harmonic approximation, the entropy cost, $\Delta S(l) = b - c\ln l$, can be computed analytically (resp. numerically) for the winding (resp. cyclization) constraint, $c_w = 1/2$ (resp. $c_c = 7/2$). We finally get the following free energy landscape statistical properties in the WD limit:

$$\beta\bar{f}(l) = \frac{A}{2}\left[\frac{\alpha^2}{l} + \sigma_0^2 l^{2H-1}\right] + c\ln l - b,$$
$$\beta^2\overline{[f(l) - \bar{f}(l)]^2} = A^2\alpha^2\sigma_0^2 l^{2H-2} \overset{\text{not.}}{=} \Lambda(l). \quad (4)$$

But the thermodynamical properties of the system are likely to depend on the correlations of the free energy landscape. From Eq. (3), one gets:

$$C(s' - s, l) = \beta^2\overline{\delta f(s', l)\delta f(s, l)} = \Lambda(l)C_H\left(\frac{s' - s}{l}\right), \quad (5)$$

where $\delta f(s, l) = f(s, l) - \bar{f}(l)$ and $C_H(y) = (|y + 1|^{2H} + |y - 1|^{2H} - 2|y|^{2H})/2$ is the correlation function of fractional Brownian motions (fBm) [15]. The results reported in Fig. 2(b) show that the scaling form (4) of the free energy rms fluctuations is well verified for weak disorder ($\sigma_0 = 0.01$) up to loop size $l \lesssim 10^3$. As shown in Fig. 2(c), the free energy correlation function decreases rather fast over a distance of order l, and then much slowly at larger distances (larger H, slower the decrease) in good agreement with the asymptotic behavior $C_H(y) \sim H(2H - 1) \times y^{2H-2}$ for $y \to \infty$ [Eq. (5)]. While the free energy fluctuations are short range correlated for $H = 1/2$, they display LRC for $H > 1/2$.

The thermodynamics of a single ''loop'' of size l embedded in a chain of length L is described by the partition function $Z(l, L) = \int_0^{L-l} \exp[-\beta f(s, l)]ds$, which accounts for all the possible locations of the loop along the chain. The equilibrium properties are determined by the free energy of the system (relatively to the unconstrained state of the chain): $\beta\mathcal{F}(l, L) = -\ln[Z(l, L)]$. The thermodynamics associated to rugged energy landscapes have been widely studied during the past decades, and have been shown to depend upon the statistics of energy fluctuations. When no correlations are present, it is the well-known random energy model (REM) that can be solved exactly [16]. This model presents a freezing phase transition separating a self-averaging ''high temperature'' (HT) phase where the ''constraint'' can explore all the possible configurations (positions) and a ''low temperature'' (LT) phase dominated by the few lowest energy minima where the constraint is likely to be localized [17]. But we have seen in Eq. (5) that the loop free energy fluctuations are LRC which may question the pertinence of the REM. In the HT/WD limit, $\beta\delta f(s, l) \ll 1$, $\forall\, s$ one gets for finite L:

$$\beta\bar{\mathcal{F}}(l) \simeq \beta\bar{f}(l) - \ln(L-l) - \frac{\beta^2}{2}\overline{\delta^2 f}(l) + \frac{1}{2}\mathbf{C}(l,L), \qquad \beta^2\overline{[\mathcal{F}(l) - \bar{\mathcal{F}}(l)]^2} = \mathbf{C}(l,L) = \frac{1}{L^2}\iint C(s-s',l)\,ds\,ds'. \quad (6)$$

The correlations control the sample-to-sample fluctuations. An explicit computation gives for both the winding and cyclization constraints: $\mathbf{C}(l,L) \sim \Lambda(l)(L/l)^{2H-2} \propto L^{2H-2}$. The correlations vanish independently of l, in the thermodynamic limit $L \to \infty$ leading to the asymptotic validity of the REM. Combining Eqs. (4) and (6), one gets in the HT/WD phase:

$$\beta\bar{\mathcal{F}}(l) = \frac{A\alpha^2}{2l} + \frac{A\sigma_0^2}{2}l^{2H-1} - \frac{A^2\alpha^2}{2}\sigma_0^2 l^{2H-2} + c\ln l - b - \ln L. \quad (7)$$

In Figs. 3(a) and 3(b) are reported the evolution of the free energy of the single loop system versus the size of the cyclization constraint for a disorder amplitude $\sigma_0 = 0.01$ (0.05) comparable to that obtained when using experimentally established structural tables [13]. The symbols correspond to exact numerical estimation of the free energy for five values of H that amount to strengthen LRC while the continuous curves correspond to the quenched free energy $\bar{\mathcal{F}}_H(l,L)$ averaged over 100 single loop chains. From both numerical and analytical results, one can extract the following main messages: (i) in the absence of disorder ($\sigma_0 = 0$), the "pure" system has a free energy that presents a minimum for a finite length $l^* = \alpha^2 A/2c$ [Eq. (8)]. This optimal length separates the enthalpic domain at small scale l characterized by a power law decrease of the free energy, and the entropic domain at large scale characterized by a logarithmic increase. (ii) When one adds some intrinsic uncorrelated disorder ($H = 1/2$), the l dependence of the free energy reduces (up to a constant) to an homogeneous pure case with a renormalized value of the bending flexibility $A_{\text{eff}} = A(1 - A\sigma_0^2)$ [13,18]. Thus there is no qualitative difference between an uncorrelated system and an ideal one, but introducing disorder decreases the free energy [Figs. 3(a) and 3(b)] and favors the formation of loop of smaller size $l^* = \alpha^2 A_{\text{eff}}/2c$. (iii) When considering LRC disorder, the system no longer behaves as a homogeneous one, but more importantly, in the small scale domain, both the free energy and the optimal length $l^*(H)$ decrease [Fig. 3(d)] when one increases H. As shown in Fig. 3(c), for a fixed loop size $l = A = 200$, the quenched average free energy provides a good description of the free energy of a typical single loop chain for both $\sigma_0 = 0.01$ and 0.05. Note that only for $\sigma_0 = 0.01$ and value of $H \lesssim 0.7$, these results are well accounted by the HT/WD approximation [Eq. (7)]. Similar results are obtained in Fig. 3(d) for the optimal loop length $l^*(H)$ which decreases down to values of about a few hundreds when increasing H from 0.5 to 0.9. For $\sigma_0 = 0.01$, the solution of the HT/WD perturbative equation:

$$(2H-1)A\sigma_0^2 l^{2H} + 2cl - (2H-2)\alpha^2 A^2 \sigma_0^2 l^{2H-1} = A\alpha^2, \quad (8)$$

provides a rather good description of the H dependence of the loop size $l^*(H)$ of a typical single loop chain. The perturbative expression of the free energy [Eq. (7)] breaks down when the energy fluctuations become too large: this is the freezing transition towards the LT/SD phase where the replica approach needs to be used to get the correct quenched free energy [17]. Note that for parameter values compatible with DNA characteristic properties, namely $A = 200$, $\alpha = 2\pi$, and $\sigma_0 \sim 0.01$, the HT/WD approximation is likely to apply (Fig. 3).

A convenient formalism to investigate diffusion process in the random 1D potential $E(s,l)$ of the single fixed length loop is that of mean first passage time (MFPT) [19]. The MFPT at the position N (starting from $s = 0$) is given by:

$$\tau(N,l) \simeq 2\int_0^N ds \int_s^N ds' \exp\{2\beta[E(s) - E(s')]\}. \quad (9)$$

The average over all possible realizations of the disordered energy landscape leads to:

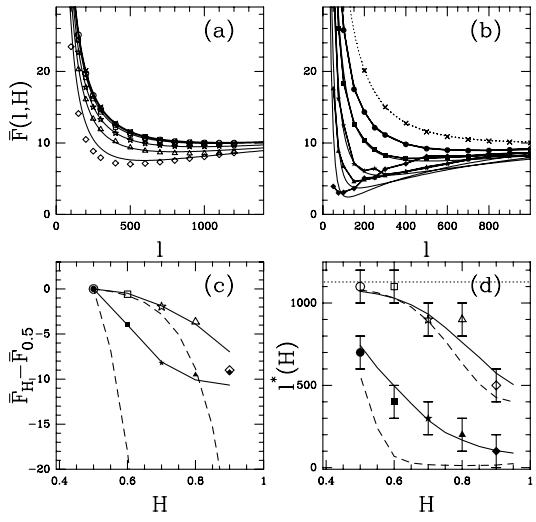

FIG. 3. Free energy of the single loop system $\mathcal{F}_H(l,L)$ vs l under the cyclization constraint and for $L = 15000$, $A = 200$, $\alpha = 2\pi$, and $\sigma_0 = 0.01$ (a) and 0.05 (b). The symbols correspond to a single chain's free energy $\mathcal{F}_H(l,L)$ obtained from the (exact) numerical computation of $f(s,l)$ for $H = 0.5$ (□, ■), 0.6 (□, ■), 0.7 (☆, ★), 0.8 (△, ▲) and 0.9 (◇, ◆); the (×) correspond to the pure case without disorder. The continuous curves stand for the corresponding quenched free energies $\bar{\mathcal{F}}_H(l,L)$ averaged over 100 "typical" single chain free energies computed using Eq. (3) for the enthalpic part (i.e., the winding energy) and $c_c = 7/2$ for the entropy cost; the dotted curve corresponds to the exact expression for the pure case. (c) $\mathcal{F}_H(l) - \mathcal{F}_{1/2}(l)$ vs H between LRC and uncorrelated chains, for loop size $l = 200$; the dashed curves correspond to perturbative approximation. (d) Optimal loop length $l^*(H)$ vs H; the dashed curves correspond to the perturbative expression (10); the horizontal line indicates the optimal loop length $l^* = 1128$ for the pure system. In (c) and (d) the symbols and the continuous curves have the same meaning as in (a) and (b).

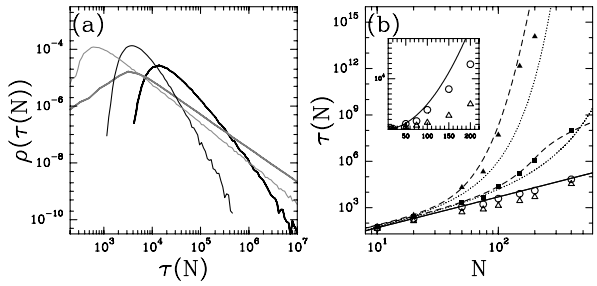

FIG. 4. MFPT $\tau(N)$ (measured in number of steps) for $A = l = 200$, $\alpha = 2\pi$, $L = 15000$, and $\sigma_0 = 0.01$. (a) PDF of $\tau(N)$ calculated for 100 000 uncorrelated ($H = 0.5$, black) and LRC ($H = 0.8$, gray) chains for $N = 100$ (thin) and 200 (thick). (b) Most probable MFPT vs N for $H = 0.5$ (○) and 0.8 (△); mean MFPT vs N for $H = 0.6$ (■) and 0.8 (▲); the dashed curves correspond to the analytical quenched average [Eq. (10)]; the dotted curves correspond to the small displacement approximation [Eq. (11)]; the continuous line corresponds to the pure diffusion case $\tau(N) = N^2$.

$$\overline{\tau}(N, l) \simeq 2 \int_0^N ds \int_s^N ds' e^{4\Lambda(l)[1 - C_H((s'-s)/l)]}. \qquad (10)$$

When looking at displacements smaller or of the order of the loop size, $N \lesssim l$, then the typical energy barrier increases like $\Delta E(N) \sim N^H$: the energy landscape has a fBm structure. For $(s' - s)/l \ll 1$, Eq. (10) reduces to:

$$\overline{\tau}(N, l) \sim N^2 e^{[4/(2H+1)(H+1)]\Lambda(l)(N/l)^{2H}}. \qquad (11)$$

We thus get a stretched exponential creep that depends on H. For $H = 1/2$, one recovers the exponential creep of the random force model with logarithmically slow (''Sinaï'') diffusion [20]. When strengthening the LRC by increasing $H > 1/2$, one further increases $\overline{\tau}(N, l)$ suggesting some slowing down of the loop dynamics. In Fig. 4(b), this modified Sinaï diffusion [20] accounts quite well for the short distance dynamics of single loop chain realizations. But as shown in Fig. 4(a), when computing the probability density function (PDF) of the MFPT for distances $N \lesssim 2l$, the way the average MFPT depends on H is very much affected by the evolution of the PDF tail and does not reflect the dependence of the most probable MFPT which, in contrast, decreases when increasing H. This shows that for a typical event, the motion of a single loop in a LRC chain over distances of the order of its size is definitely superdiffusive, larger H, faster the dynamics.

To summarize, we have shown that the competing effects of entropy and sequence-dependent structural disorder favor the autonomous formation of DNA loops. When taking into account the existence of LRC as observed in eukaryotic genomic sequences [12], we have found, in the WD limit, that strengthening LRC allows the formation of smaller 2D loops that superdiffuse, larger the LRC, faster the typical local loop dynamics. These results strongly suggest that these LRC predispose eukaryotic DNA to interact with histones to form nucleosomes. The size of the selected loops (few hundreds bp) are typical of the characteristic DNA which is wrapped around histones. The local rapid diffusion of the loop induced by the LRC structural disorder provides a very attractive interpretation to the nucleosome repositioning dynamics. LRC are likely to help the nucleosomes to rearrange themselves in a very efficient way as, e.g., after the passage of the transcription and replication polymerases. Since in chromatin, the nucleosomal string presents a high occupation density with an average distance between nucleosomes of the order of 50 bp, this raises the issue of the effect of the interaction between nucleosomes on their large scale mobility. The generalization of the present work to multiple 2D loops in a long LRC DNA chain is in current progress.

[1] A. P. Wolffe, *Chromatin Structure and Function* (Academic, New York, 1998), 3rd ed..
[2] J. Widom, Annu. Rev. Biophys. Biomol. Struct. **27**, 285 (1998); R. D. Kornberg and Y. Lorch, Cell **98**, 285 (1999).
[3] K. Luger *et al.*, Nature (London) **389**, 251 (1997).
[4] P. B. Becker, EMBO J. **21**, 4749 (2002).
[5] G. Meersseman, S. Pennings, and E. M. Bradbury, EMBO J. **11**, 2951 (1992); A. Flaus and T. J. Richmond, J. Mol. Biol. **275**, 427 (1998).
[6] H. Schiessel, J. Phys. Condens. Matter **15**, R699 (2003).
[7] H. Schiessel, J. Widom, R. F. Bruinsma, and W. M. Gelbart, Phys. Rev. Lett. **86**, 4414 (2001); **88**, 129902 (2002); I. Kulić and H. Schiessel, Biophys. J. **84**, 3197 (2003).
[8] F. Mohammad-Rafice, I. M. Kulić, and H. Schiessel, J. Mol. Biol. **344**, 47 (2004).
[9] J. C. Marini *et al.*, Cold Spring Harbor Symposia on Quantitative Biology (1933-) [Proceedings] **47**, 279 (1983).
[10] I. Ioshikes *et al.*, J. Mol. Biol. **262**, 129 (1996); A. Thaström *et al.*, J. Mol. Biol. **288**, 213 (1999).
[11] T. E. Cloutier and J. Widom, Mol. Cell **14**, 355 (2004).
[12] B. Audit *et al.*, Phys. Rev. Lett. **86**, 2471 (2001); J. Mol. Biol. **316**, 903 (2002).
[13] C. Vaillant, B. Audit, C. Thermes, and A. Arneodo, Phys. Rev. E **67**, 032901 (2003).
[14] Y. Zhang and D. M. Crothers, Biophys. J. **84**, 136 (2003).
[15] B. B. Mandelbrot and W. van Ness, SIAM Rev. **10**, 422 (1968).
[16] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[17] J.-P. Bouchaud and M. Mézard, J. Phys. A **30**, 7997 (1997).
[18] P. Nelson, Phys. Rev. Lett. **80**, 5810 (1998).
[19] M. Slutsky, M. Kardar, and L. A. Mirny, Phys. Rev. E **69**, 061903 (2004).
[20] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990); T. Hwa *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).